A NEW OUTCOME MEASURE FOR COST-UTILITY ANALYSES OF SCREENING PROGRAMS*

Carmen Herrero and Juan D. Moreno-Ternero**

WP-AD 2003-24

Corresponding author: Carmen Herrero, Departamento de Fundamentos del Análisis Económico. Universidad de Alicante. 03071 Alicante. Spain. Phone: (34)965903618. Fax (34)965903685. E-mail: carmen.herrero@ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Julio 2003
Depósito Legal: V-3474-2003

* We are most grateful to Han Bleichrodt for detailed comments and suggestions. Thanks are also due to Mike Drummond, Dunia López-Pintado, Ignacio Ortuño-Ortin and Lars Østerdal for helpful discussions. Part of this research was conducted while the second author was visiting the Institute of Economics at the University of Copenhagen, thanks to a Marie Curie Fellowship. Financial support from the Ministerio de Ciencia y Tecnología, under project BEC2001-0535, and from the Generalitat Valenciana under project CTIDIB/2002/314, is gratefully acknowledged.

** C. Herrero and J.D. Moreno: Departamento de Fundamentos del Análisis Económico, Universidad de Alicante.

A new outcome measure for cost-utility analyses of screening programs

Carmen Herrero and Juan D. Moreno-Ternero

ABSTRACT

In this paper we provide a new outcome measure for the cost-utility analyses of alternative screening programs of a particular disease. We show that for non-invasive screening programs satisfying plausible assumptions, QALYs can be replaced by a simpler outcome: the sensitivity of the program. In other words, the cost-utility analysis can be made without computing the utility each program offers. Consequently, results would be immune to two of the most controversial issues in the cost-utility analysis approach: the elicitation method to obtain quality weights of health profiles, and the discount rate for future health benefits. The assumptions are particularly suitable in the case of selecting between the universal and the selective implementation of a non-invasive screening program. Therefore, we apply our result to provide an additional viewpoint in the current debate about the implementation of a universal or selective newborn screening program to detect congenital hearing impairment.

JEL classification numbers: I10

Key words: Cost-utility analysis, cost-sensitivity ratios, screening programs, newborn hearing screening.

1 Introduction

In health care, as in other areas of social policy, decisions have to be made concerning the allocation of scarce resources. The economic evaluation of the available alternative health care programs helps in guiding this resources allocation. One of the most frequently used methods for the economic evaluation of health care programs is the *Cost Utility Analysis* ("*CUA*" hereafter). CUA measures the benefits of a program in utility terms. The utility index that has been most frequently employed is the *Quality Adjusted Life Years* ("*QALYs*" hereafter). On the basis of the QALY index, preferences on alternative programs can be given by means of cost-utility (or "cost per QALY gained") ratios. In particular, the program that should be implemented, among a set of alternative ones, is the one that offers the largest number of QALYs per dollar or, what is equivalent, the one that has the lowest cost per QALY gained.

An interesting class of health care programs is that of screening programs, the one we consider in this paper. Screening is traditionally defined as testing a population of asymptomatic individuals to identify precursors of a disease. The subjects who test positive are sent on for further evaluation in a subsequent diagnostic test to determine whether they do, in fact, have the disease. An implicit assumption underlying the clinical interest of screening programs is that early detection, before the development of symptoms, will lead to a more favorable prognosis. This is so because, by means of a screening, it is possible to treat the disease before it becomes clinically manifest, which is more effective than a later treatment.

The validity of a screening test is defined as its ability to distinguish between those who have the disease and those who do not. Validity has two components: *sensitivity* and *specificity*. The *sensitivity* of the test is defined as the ability of the test to correctly identify those who *have* the disease. The *specificity* of the test is defined as its ability to correctly identify those who *do not have* the disease. Obviously, the best scenario is to have a screening test that is both highly sensitive and highly specific. This is not usually feasible, and, in general, there is a trade-off between the sensitivity and specificity of any given screening test.

Typical decision problems that arise when facing a preventable disease for which there exist screening programs are the following: *(1) Is it worth to implement a screening program for the early detection of the disease?; (2) If so, which of the available programs should be chosen?; (3) Would it be more effective to implement a screening program to the whole population or just to those belonging to a group at risk?*

In this work, we focus on the last two questions. More precisely, the

aim of this paper is to show that, under appropriate and mild assumptions, the CUA of alternative screening programs of a particular disease can be made without computing the QALYs each program offers. More precisely, under such assumptions, QALYs can be replaced by a simpler outcome: the sensitivity of the program. Consequently, the screening program that should be implemented, in accordance with a CUA, is the one that shows the lowest cost-sensitivity ratio.

In particular, our result shows that the CUA of screening programs of a particular disease, satisfying the assumptions, is immune to two of the main criticisms to the CUA approach. The first one is the lack of consensus over the method to be used to elicit health state utility weights.[1] The second one is the controversy in the selection of a discount rate to be used to obtain present values of future health benefits.[2]

The assumptions leading to our result can be simplified to solve question (3) above, i.e., in the case of deciding between applying a screening program to the whole population (a universal alternative) or just to a subset of it (a selective alternative). An example where to apply our result is precisely the current debate on the implementation of newborn hearing screening programs in some states of the United States and in some European countries (e.g. [9], [10], [11], [12]).

The rest of the paper is organized as follows. First, we set up the preliminaries of our model. Second, we provide the main result and its simplification to the case of universal versus selective alternatives. Then, we apply the new technique to the case of newborn hearing screening procedures. Finally, we present the conclusions. The proofs of the results have been relegated to an Appendix.

[1] There are three standard elicitation methods: rating scale, time trade-off, and standard gamble (for a description of these three methods see, for instance, [1]). These methods lead, systematically, to different weights (e.g. [2], [3], [4], [5], [6]). As a result, different health care programs might emerge as the preferred ones, depending on the method chosen.

[2] There is broad agreement that, in CUA, individuals generally prefer health gains to occur earlier than later. There is no agreement, however, about how to convert future health consequences, expressed in utility terms, in present values. One possibility to deal with this task is the introduction of a discount rate. Different recommendations have been made for the choice of an appropriate discount rate (e.g. [1], [6]). Furthermore, there is also controversy about whether the constant rate discounted utility model should be used at all (c.f. [7], [8]).

2 Preliminaries

We consider a particular preventable disease d. Typical examples of d could be breast cancer, congenital hearing impairment, or even heart diseases. Let $\Gamma = \{1, ..., n\}$ be the corresponding target population of agents susceptible of suffering the disease d. For instance, if d refers to a congenital disease then Γ typically refers to a cohort of newborns. We denote by $p \in [0, 1]$ the prevalence of d in the target population. That is, if i denotes the number of impaired individuals, then $p = \frac{i}{n}$. Individual status with respect to d is either *positive* (if the agent is impaired) or *negative* (if the agent is healthy). We denote by G_d the set of individuals with positive status with respect to d, and by G_{nd} the set of individuals with negative status with respect to d. By construction, $\Gamma = G_d \cup G_{nd}$ and the number of individuals in each of the subgroups is $|G_d| = p \cdot n$ and $|G_{nd}| = (1 - p) \cdot n$, respectively.

Let us suppose that there exists a set $S = \{s_1, ..., s_m\}$ of available screening programs for the early detection of d. For the sake of completeness, we denote by s_0 the 'status quo', i.e., the scenario with no screening program. Each screening is determined by three parameters: cost, sensitivity and specificity. Formally, for all $j = 0, 1, ..., m$, $s_j = s_j(c_j, se_j, sp_j)$, where c_j denotes the cost, se_j the sensitivity and sp_j the specificity of s_j.

Table 1: Impairment status and test results

Status (S) / Test (T)	Positive (S^+)	Negative (S^-)	Total
Positive (T^+)	TP	FP	$TP + FP$
Negative (T^-)	FN	TN	$FN + TN$
Total	$TP + FN$	$FP + TN$	n

Prevalence: $p = \frac{TP+FN}{n} = \frac{TP+FN}{TP+FP+FN+TN}$
Specificity: $sp = P(T^-|S^-) = \frac{TN}{FP+TN}$
Sensitivity: $se = P(T^+|S^+) = \frac{TP}{TP+FN}$
True positives: $TP = p \cdot se \cdot n$
False positives: $FP = (1 - p) \cdot (1 - sp) \cdot n$
False negatives: $FN = p \cdot (1 - se) \cdot n$
True negatives $TN = (1 - p) \cdot sp \cdot n$

Since the sensitivity and specificity of a screening test have been described in the introduction, we shall not dwell on their definitions here (see Table 1 for further details). However, the concept of 'costs of a screening program'

deserves some further explanations. First, for ease of exposition, we consider *costs per capita* rather than absolute costs. Second, by the costs of a screening program we mean all sort of health care expenses associated with the program. That is, the costs of the screening technique, and the costs of the final diagnostic test to which every individual that has been identified as positive by the screening is referred. Thus, note that the sensitivity and the specificity of a program influence its costs.

After the implementation of a screening program, there is a subsequent population partition into four groups, depending on the test results and the disease status. More precisely, for each $s_j \in S$ let us denote by G^j_{TP} the group of true positive individuals after implementing s_j. Similarly, G^j_{FP} (G^j_{FN}) [G^j_{TN}] denotes the group of false positive (false negative) [true negative] individuals after implementing s_j. Then, for each $s_j \in S$, Γ is expressed as follows:

$$\Gamma = G^j_{TP} \cup G^j_{FP} \cup G^j_{FN} \cup G^j_{TN}. \tag{1}$$

According to the notation introduced above, it is straightforward to see that the probabilities of being in each of the groups are given by:

$$\begin{aligned} p^j_{TP} &= p \cdot se_j \\ p^j_{FP} &= (1-p) \cdot (1-sp_j) \\ p^j_{FN} &= p \cdot (1-se_j) \\ p^j_{TN} &= (1-p) \cdot sp_j \end{aligned}$$

Finally, we define a concept that will be useful in the ensuing discussion: the *cost-sensitivity ratio* ("*CS-ratio*" hereafter). The CS-ratio of a screening program is the ratio between its cost and its sensitivity, i.e., the cost per level of sensitivity that the program offers. More precisely:

Definition 1 *Given $s_j \in S$, its CS-ratio is given by $R_j = \frac{c_j}{se_j}$.*

Note that the specificity of a screening program is also contemplated in its CS-ratio, since, as we mentioned above, it is necessary to compute the costs of the program.

3 The main result

The aim of this section is to select the best screening program among those in the set S, according to a CUA. In order to do that, let Q be the QALY index that gives for each individual health profile its number of QALYs. Note

that Q depends on the particular elicitation method and discount rate chosen. Thus, strictly speaking, $Q(\cdot) = Q^{e,\delta}(\cdot)$, where e denotes the elicitation method and δ the discount rate. In what follows, and for ease of notation, we avoid superscripts in the ensuing discussion when referring to index Q.

According to the notation in Section 2, the expected number of QALYs for an individual, after implementing a program s_j, is given by:

$$Q_j = p^j_{TP} \cdot Q^j_{TP} + p^j_{FP} \cdot Q^j_{FP} + p^j_{FN} \cdot Q^j_{FN} + p^j_{TN} \cdot Q^j_{TN},$$

where Q^j_{TP}, Q^j_{FP}, Q^j_{FN} and Q^j_{TN} are the corresponding expected number of QALYs for an individual in each of the resulting groups in the population partition (1). Similarly, the expected number of QALYs for an individual in the 'status quo' is given by:

$$Q_0 = p \cdot Q_d + (1-p) \cdot Q_{nd},$$

where p is the prevalence of d and Q_d and Q_{nd} are the expected number of QALYs for an individual in the groups G_d and G_{nd} respectively.

Now, since $c_0 = 0$, the cost-utility (or cost-per-QALY-gained) ratio that corresponds to the screening program s_j is given by $c_j/(Q_j - Q_0)$. CUA provides preferences among the alternative options by means of such ratios. In particular, the preferred program would be the one that exhibits the lowest cost-utility ratio, i.e.,

$$s^* = \arg\min \left\{ \frac{c_j}{Q_j - Q_0} : s_j \in S \right\}.$$

Furthermore, we can establish a ranking order among all the alternative options according to the ratios. More precisely, let us define

$$\begin{aligned} s^{(1)} &= s^*, \\ s^{(l)} &= \arg\min\{\frac{c_j}{Q_j - Q_0} : s_j \in S \setminus \{s^{(1)}, ..., s^{(l-1)}\}\}, \text{ for all } l = 2, ..., m-1, \\ s^{(m)} &= \arg\max\{\frac{c_j}{Q_j - Q_0} : s_j \in S\}. \end{aligned}$$

Thus, under the decision framework of a CUA, preferences among the alternative options would be the following:

$$s^{(1)} \succsim s^{(2)} \succsim ... \succsim s^{(m)}.$$

Now, we show that, under some additional assumptions, the CUA that has just been described can be reduced to the analysis of the CS-ratios introduced above.

The first assumption says, roughly, that utility does not decrease 'per se' by being referred to a screening program. In other words, the number of QALYs of a true (false) negative individual after implementing a screening program coincides with the number of QALYs of a healthy (impaired) individual in the status quo. The second assumption says that early detection of the disease is advantageous at an individual level, and that this individual improvement is independent of the screening method chosen. The third assumption says that there are no utility differences between healthy individuals with different test results, i.e., between a false positive and a true negative individual.

The plausibility of these assumptions depends on the screening programs we are considering. They are fulfilled whenever the alternative programs are non-invasive and such that false positives are correctly identified in a short period of time. In the case of newborn screening programs, the three assumptions are sound.

Formally, the assumptions are the following:

Assumption 1: *For each QALY index Q, and for all $s_j \in S$, we have*

$$Q^j_{TN} = Q_{nd}, \text{ and } Q^j_{FN} = Q_d.$$

Assumption 2: *For each QALY index Q, and for all $s_j \in S$, there exists $\gamma_Q > 0$ such that*

$$Q^j_{TP} - Q^j_{FN} = \gamma_Q.$$

Assumption 3: *For each QALY index Q, and for all $s_j \in S$, we have*

$$Q^j_{FP} = Q^j_{TN}.$$

The main result of the paper comes next.

Theorem 1 *Under Assumptions 1 to 3, the screening program for a particular disease that should be implemented according to a CUA, is the one that shows the lowest CS-ratio.*

(The proof can be found in the Appendix)

From the proof of Theorem 1, it can be inferred that preferences among the set of available screening programs, according to a CUA, can be obtained from the CS-ratios, whenever Assumptions 1 to 3 are fulfilled. In other words, if that is the case, both the cost-utility ratios and the CS-ratios yield the same ranking of preferences for the set of alternative programs. In this case, our result permits us to circumvent two of the main sources of controversy in a

CUA: the choice of an elicitation method (or a generic utility measure) and the choice of a discount rate.

Roughly speaking, by Theorem 1 we can perform a CUA of screening programs relying on information limited to the intrinsic properties of each program: cost, sensitivity and specificity. The use of the statistical properties of the test resembles the main method in the literature on measuring the accuracy of diagnostic systems: the so-called ROC analysis (e.g. [13], [14]). A ROC analysis is a method of assessing the value of a screening (or a diagnostic) test by deciding where to put the threshold when using the test. Where to set the threshold requires consideration on two factors: the total number of errors (in general, errors of type I, or false negative cases, and errors of type II, or false positive cases) made, and the importance of errors of type I versus errors of type II. The threshold should be chosen so that we minimize the total error rate, i.e., the weighted aggregation of type I errors and type II errors, where weights are chosen so that they reflect the importance of each type error. Our Assumption 3 implies that the importance of false positives is negligible. Thus, a ROC analysis of screening programs satisfying our assumptions would also rank them according to their sensitivity levels. This is precisely what we obtained from Theorem 1 in the case of ranking these programs according to the QALYs they offer. This adds robustness to our main result.

4 Universal vs. selective screening programs

In this section we face a related problem to the one in the previous section. We assume that a particular screening technique has shown to be superior to the other available ones and the issue is to decide whether it would be more effective to implement the screening for the whole population, i.e., a *universal* alternative, or just for the group of individuals with risk factors for the disease, i.e., a *selective* alternative. The optimal decision will obviously depend on the portion of impaired individuals that are not included in the group of individuals with risk factors. The higher this portion, the higher the number of impaired individuals that are not detected by a selective alternative and therefore, the lower the effectiveness of this option.

We could tackle this issue by applying our main result. To do so, one just need to consider $S = \{s_1, s_2\}$ where s_1 is the universal alternative (for which we know the parameters, namely, sensitivity, specificity and costs, that determine it) and s_2 is the selective alternative (whose parameters can be obtained from those of s_1 and the prevalence of the disease in the group of individual with risk factors). Thus, under the assumptions of our main

result, the CUA to select among these two alternatives can also be made without computing the QALYs each one offers.

Now, we show in this section that our model can be simplified for this particular framework, as follows. Let $s = s(se, sp, c)$ be a particular screening program for the early detection of d. Let $\widehat{s}$ be the same program but only applied to the population $\Gamma^R \subset \Gamma$ of individual with risk factors of suffering d. Let r be the probability of being in Γ^R, p_r the prevalence of the disease in Γ^R and $\widehat{c}$ the cost (per capita) of implementing $\widehat{s}$. Let Q be a QALY index. We denote by Q^s and $Q^{\widehat{s}}$ the expected number of QALYs for an individual after implementing s and $\widehat{s}$ respectively. According to the notation in Section 2,

$$Q^s = p \cdot se \cdot Q^s_{TP} + (1-p) \cdot (1-sp) \cdot Q^s_{FP} + p \cdot (1-se) \cdot Q^s_{FN} + (1-p) \cdot sp \cdot Q^s_{TN}.$$

Furthermore, it is straightforward to show that

$$\begin{aligned} Q^{\widehat{s}} \;=\; & r \cdot p_r \cdot se \cdot Q^s_{TP} + r \cdot (1-p_r) \cdot (1-sp) \cdot Q^s_{FP} + \\ & r \cdot p_r \cdot (1-se) \cdot Q^s_{FN} + r \cdot (1-p_r) \cdot sp \cdot Q^s_{TN} + \\ & (p - r \cdot p_r) \cdot Q_d + [(1-r) - (p - r \cdot p_r)] \cdot Q_{nd}. \end{aligned}$$

We now adapt the assumptions of our main result to this new framework.

The first assumption says that early detection of the disease by means of the screening program is advantageous at an individual level. The second assumption says that there are no utility differences between healthy individuals with different results in the screening. Furthermore, the assumptions say that individual utility does not decrease by being referred to the screening program. Formally:

Assumption 1*: *For each QALY index Q, we have*

$$Q_d = Q^s_{FN} < Q^s_{TP}.$$

Assumption 2*: *For each QALY index Q, we have*

$$Q_{nd} = Q^s_{FP} = Q^s_{TN}.$$

For the sake of completeness, before stating the result of this section, we introduce a piece of notation. We call the *cost per impaired individual* associated to a screening program to the ratio between its cost and the number of impaired individuals it eventually identifies. Formally, the cost per impaired individual associated to s is

$$\frac{c \cdot n}{p \cdot n} = \frac{c}{p},$$

whereas the cost per impaired individual associated to $\widehat{s}$ is

$$\frac{\widehat{c} \cdot n}{r \cdot p_r \cdot n} = \frac{\widehat{c}}{r \cdot p_r}.$$

The following result is obtained:

Theorem 2 *Under Assumptions 1* and 2*, s is CUA superior to $\widehat{s}$ if and only if the cost per impaired individual associated to s is lower than the cost per impaired individual associated to $\widehat{s}$.*

(The proof can be found in the Appendix)

Theorem 2 says that we can decide between a universal and a selective alternative, according to a CUA, just with the information of the parameters that determine the screening program. Note that the sensitivity and the specificity of the program are also captured in the cost per impaired individual associated to each alternative, as they determine the number of individuals referred to the diagnostic test.

5 Application: the case of congenital hearing impairment

We conclude by applying our model to the case of congenital hearing impairment. The hearing impairment satisfies all the medical requirements to impose a prevention program, based on a newborn screening protocol. First of all, it is a serious disease, for which a lack of early diagnosis will cause problems in language acquisition. Significant hearing loss interferes with the development of speech perception abilities needed for later language learning. These impairments in communication skills can lead to learning disabilities and ultimately, to limitations in career opportunities. Moreover, it is more frequent than other impairments for which newborn screening programs are in use in developed countries. Finally, there are reliable screening methods, with high levels of sensitivity and specificity, and there is also an effective treatment available.

Due to these facts, there is a broad agreement to impose a newborn hearing screening program, as subsequently recommended the *1993 National Institutes of Health Consensus statement on the early identification of hearing impairment in infants and young children* [9], the *1999 European consensus statement on neonatal hearing screening* [10] and the *Year 2000 position statement* of the *Joint Committee on Infant Hearing* [11].

Having reached this consensus, the debate moved to select between a universal and a selective alternative. In a *Universal Newborn Hearing Screening* (*"UNHS"* hereafter) every newborn is tested, whereas in a *Selective screening* only those who were born with a risk factor, such as being in the neonatal intensive care unit or having a family history of hearing impairment, are tested. A UNHS is more expensive but also more effective, since only 50% of newborns with a hearing impairment belong to a group at risk [9]. It is currently mandated in 32 states of the United States [12]. The selective screening, however, was and continues to be practiced throughout the United States and the rest of the world [12].

There is ample literature on choosing among UNHS and selective screening, especially from the medical viewpoint (see [9], [10], [11], [15], [16] and [17] among others), but also from an economic viewpoint (see [12], [18], [19], [20] and [21] among others). In the medical literature, the debate seems to be closed, in favor of the universal procedure [11]. Every neonate should be tested by *Automated Transient Evoked Otoacoustic Emissions* (*"TEOAE"* hereafter), a less efficient and expensive test, followed by *Automated Auditory Brainstem Responses* (*"AABR"* hereafter), a more efficient and expensive test, for those who failed the initial stage [11]. The aim of this section is to apply our model to provide an additional economic viewpoint to this current debate about choosing between the two alternatives. As we will see later, our conclusions do not fully agree with the recommendations of the Joint Committee on Infant Hearing.

5.1 Data

We conduct a CUA of the universal protocol and the selective protocol from the hospital's viewpoint. That is, we only compute direct costs associated to them.[3] Keren et al. [12] is the most recent and comprehensive source of data that exists for this problem. We consider a hypothetical birth cohort of 80000 infants in one state as they do. The decision model compares the two protocols. The universal protocol applies the 2-stage process mentioned above to every newborn in such a cohort. The selective screening includes a previous stage with a high-risk criterion, and then applies the same 2-stage process for infants in the cohort at risk for congenital hearing loss. Each protocol concludes with a diagnostic test for those who failed after the last stage.

Table 2 shows the estimates of the probabilities in the decision model.

[3]Other viewpoints that consider indirect costs, like special education costs, long-term productivity or disability allowances, are considered in [12], [20] and [21].

Additional information about them can be obtained in [12], [18], [19] or [20].

Table 2: Probabilities

Variable	Mean	Range	References
Proportion of newborns at risk (r)	0.13	$0.08 - 0.15$	[12], [18]
Prevalence (p)	0.0016	$0.001 - 0.006$	[12], [19]
Prevalence within newborns at risk (p_r)	0.008	$0.001 - 0.05$	[12], [18]
Sensitivity (se)	0.903	$0.81 - 1$	[12], [19]
Specificity (sp)	0.985	$0.93 - 1$	[12], [19]

From this table, we can easily obtain the mean probability of being referred to the diagnostic test by each protocol. Formally,

$$p^+ = p_{TP} + p_{FP} = p \cdot se + (1 - p) \cdot (1 - sp) = 0.016,$$

and

$$\widehat{p}^+ = \widehat{p}_{TP} + \widehat{p}_{FP} = r \cdot p_r \cdot se + r \cdot (1 - p_r) \cdot (1 - sp) = 0.003.$$

Table 3 shows the mean costs in the decision model. All costs were adjusted to 2001 US dollars. Future costs were discounted at a rate of 3% per year, as recommended by the Panel on Cost-Effectiveness in Health and Medicine [6].

Table 3: Costs

Variable	Universal	Selective	References
Screening technique cost	1555200	777800	[12], [20]
Screening technique cost (per infant) (c_s, $\widehat{c_s}$)	19	10	[12], [20]
Diagnostic evaluation (per infant) (c_d)	540	540	[12]
Risk factor detection (per infant) (c_r)	–	0.95	[12], [18]
Screening cost (per infant) (c, $\widehat{c}$)	28	11	[20]

Screening costs encompass the screening technique, i.e., machines, supplies and wages, and the final diagnostic test to which every infant that has been identified as positive by the screening is referred.[4] Formally, if c denotes the cost (per infant) of the screening, then

$$c = c_s + p^+ \cdot c_d,$$

where c_s is the cost (per infant) of the screening technique, c_d is the cost (per infant) of the diagnostic test, which is independent of the screening program,

[4]For further details about them see [12] or [20].

and p^+ is the probability of being identified as positive by the screening. Similarly, $\widehat{c}$, the cost (per infant) of implementing $\widehat{s}$, is

$$\widehat{c} = \widehat{c_s} + \widehat{p^+} \cdot c_d + c_r,$$

where $\widehat{c_s}$ is the cost (per infant) of the screening technique, $\widehat{p^+}$ is the probability of being identified as positive by $\widehat{s}$, and c_r is the cost of detecting a risk factor to one infant.

5.2 Results

The selective protocol is less expensive but the UNHS produces higher utility gains. Thus, according to a CUA, we need the 'cost-per-QALY-gained' ratios to decide. Thanks to Theorem 2, and if Assumptions 1^* and 2^* are fulfilled, we can decide which protocol is CUA superior, without computing the QALYs they offer. In particular, for any elicitation method and discount rate chosen, the conclusions of the CUA would be the same.

Now, in the framework of newborn hearing screening, these assumptions are mild. For instance, recent works have played down the importance of the main impact of false positives -the hypothetical anxiety or depression of their parents- in the early detection of congenital hearing impairment (e.g. [22], [23], [24], [25]). Furthermore, given that the screening protocol is not very invasive (e.g. [9], [11]) we can accept that individual utility does not decrease by being referred to the screening.

From Tables 2 and 3 we obtain the cost per impaired individual associated to each protocol:

$$\frac{c}{p} = 17500 > 10600 = \frac{\widehat{c}}{r \cdot p_r}.$$

Thus, the cost per impaired individual associated to the selective protocol is lower, which shows that the selective screening is CUA superior to the UNHS, according to Theorem 2.

The univariate sensitivity analyses tell us that the uncertainty over the cost of detecting a high risk factor in a newborn has substantial influence, due to its wide confidence interval. More precisely, although the mean estimate of the cost of detecting a high-risk factor in a newborn is \$0.95 as Table 3 shows, the confidence interval for such a parameter is $[\$0.5, \$15]$ [18]. It can be shown that, if the value of the parameter varies across its confidence interval, the conclusion would not hold. More precisely, if it would increase from its mean estimation \$0.95, to \$7.5 then UNHS would lead to a lower cost per impaired individual. Other variables to which the model is moderately

sensitive are the prevalence (p), the prevalence within newborns at risk (p_r) and the proportion of newborns at risk (r) (see [20] for further insights).

Table 4: Cost-utility ratios

	CU-RATIOS ($/QALY)			
Discount rate	$\delta = 0$		$\delta = 0.05$	
Generic utility measure	EQ-5D	HUI3	EQ-5D	HUI3
Selective screening	**675**	**440**	**3105**	**1875**
UNHS	1125	735	5170	3120

For the sake of completeness, in Table 4 we show the cost-utility ratios provided by two classical generic utility measures such as the *EuroQol 5 dimensions (EQ-5D)* and the *Health Utility Index (HUI3)*. We first show the ratios without discounting, and then with a discount rate of 0.05. Such ratios yield the cost per QALY gained, depending on the generic utility measure and the discount rate chosen.[5] The cost per QALY gained is an interesting information per se, whose main advantage is that comparisons with other programs can be made. The difference between the numbers we show in the table is that all the ratios that are obtained making use of the EQ-5D measure are higher than the corresponding ones that are obtained making use of the HUI3. It is more expensive to gain a QALY if we measure benefits with the EQ-5D measure. This is due to the fact that the HUI3 measure yields higher gains of utility for our particular problem (see [20] for the details).

To conclude, it is worth noting that all ratios reported are well below $20000 per QALY gained, the lowest critical value to adopt a new technology, according to the guidelines for economic evaluations [26]. This implies that it is worthwhile to implement a newborn hearing screening program anyhow.

6 Discussion

We have presented in this paper a new outcome measure that simplifies the QALY index for the CUA of screening programs for a particular disease. We have shown that, under some assumptions, the CUA of these programs can be made without computing the QALYs they offer and the only information that is required is the cost, the sensitivity and the specificity of each one.

The plausibility of the assumptions is clear in the case of newborn and non-invasive screening programs for a congenital disease (and even clearer

[5] We assigned utility weights according to these two generic measures by asking specialists. See [20] for further details.

when deciding between the universal and the selective implementation of a given newborn and non-invasive screening program). For these programs, the possible differences in utility between a false positive individual and a true negative one, can only lie in the hypothetical depression caused by a wrong positive test. However, and since we only compute patient's utility gains, a newborn does not get depressed over an erroneous diagnosis. One could argue that only computing patient's utility gains we would be bypassing an important negative externality, noting that the hypothetical anxiety or depression provoked by a false positive test could be translated to patient's parents. Nevertheless, if the hypothetical parental anxiety would be considered as a negative externality, then we should also take into account the positive externality caused by the care that the patient obtains from his parents, which could help considerably to his treatment. If both externalities were considered the computation of health benefits associated to these programs would not be tractable.

Furthermore, in the particular case of newborn hearing screening programs, recent studies have played down the importance of the 'false positive effect' on patient's parents. This is one of the reasons why we have applied our result to the current debate concerning the implementation of a selective or universal newborn hearing screening program. We found that the former one is preferred from the hospital's viewpoint. It is interesting to note that this conclusion differs from the preferred approach by policy makers. Under our personal opinion, this shows the value of economic evaluation of health care programs. Notwithstanding, other viewpoints computing indirect costs, like special education or disability allowances, might alter the conclusions (e.g. [20], [21]).

To conclude, we acknowledge that the technique provided in this paper is only informative when it has been decided that a screening program for a particular disease has to be implemented. It cannot determine whether it is worthwhile to have a screening program anyhow. This is a major issue in CUA since no clear decision rule exists (e.g. [26], [27]). Some analysts have suggested setting a threshold value for the cost per QALY that represents the willingness of society to pay for additional QALYs. If we adopt this rule, we could still use our technique to simplify the economic evaluation of screening programs for a particular disease. Instead of computing the 'cost-per QALY-gained' ratios of each available program, we could select the best one among them with our technique and then compare its 'cost-per QALY-gained' ratio with the threshold, to decide whether a program should be implemented or not. Therefore, we would need to compute the QALYs offered by just one program instead of doing this for each program.

7 Appendix

Theorem 1 *Under Assumptions 1 to 3, the screening program for a particular disease that should be implemented according to a CUA, is the one that shows the lowest CS-ratio.*

Proof.

Let d be a particular disease and $S = \{s_1, ..., s_m\}$ be the set of available screening programs for the early detection of d. Given $s_j \in S$ and a QALY index Q, the expected number of QALYs for an individual after implementing a program s_j, is given by:

$$Q_j = p^j_{TP} \cdot Q^j_{TP} + p^j_{FP} \cdot Q^j_{FP} + p^j_{FN} \cdot Q^j_{FN} + p^j_{TN} \cdot Q^j_{TN},$$

where Q^j_{TP}, Q^j_{FP}, Q^j_{FN} and Q^j_{TN} are the corresponding expected number of QALYs for an individual in each of the resulting groups in the population partition (1).

Upon replacing the expression of the probabilities, we have:

$$Q_j = p \cdot se_j \cdot Q^j_{TP} + (1-p) \cdot (1-sp_j) \cdot Q^j_{FP} + p \cdot (1-se_j) \cdot Q^j_{FN} + (1-p) \cdot sp_j \cdot Q^j_{TN}.$$

If s_0 denotes the 'status quo', i.e., no screening program, then the expected number of QALYs in the 'status quo' for an individual is:

$$Q_0 = p \cdot Q_d + (1-p) \cdot Q_{nd},$$

where Q_d and Q_{nd} are the expected number of QALYs for an individual in the groups G_d and G_{nd}, respectively.

Hence, the expected utility gains after implementing s_j can be expressed as

$$\begin{aligned} Q_j - Q_0 \;=\; & p \cdot se_j \cdot Q^j_{TP} + (1-p) \cdot (1-sp_j) \cdot Q^j_{FP} + \\ & p \cdot (1-se_j) \cdot Q^j_{FN} + (1-p) \cdot sp_j \cdot Q^j_{TN} - \\ & p \cdot Q_d - (1-p) \cdot Q_{nd}, \end{aligned}$$

or equivalently,

$$Q_j - Q_0 = p \cdot (se_j \cdot (Q^j_{TP} - Q^j_{FN}) + (Q^j_{FN} - Q_d)) + (1-p) \cdot (sp_j \cdot (Q^j_{TN} - Q^j_{FP}) + (Q^j_{FP} - Q_{nd})).$$

By Assumption 1,

$$Q_j - Q_0 = p \cdot se_j \cdot (Q^j_{TP} - Q^j_{FN}) + (1-p) \cdot sp_j \cdot (Q^j_{TN} - Q^j_{FP}).$$

By Assumption 3,

$$Q_j - Q_0 = p \cdot se_j \cdot (Q^j_{TP} - Q^j_{FN}).$$

Finally, by Assumption 2,

$$Q_j - Q_0 = p \cdot \gamma_Q \cdot se_j,$$

where $\gamma_Q = Q^j_{TP} - Q^j_{FN} > 0$.

Thus, for each $s_j \in S$, its cost-utility ratio is given by:

$$\frac{c_j}{Q_j - Q_0} = k \cdot R_j,$$

where

$$k = k(p, Q) = \frac{1}{p \cdot \gamma_Q} > 0,$$

and $R_j = \frac{c_j}{se_j}$ is what we called the CS-ratio of s_j.

Note that k depends on the prevalence of the disease (p), and the QALY index (Q). Therefore, it also depends on the chosen discount rate (δ), and the elicitation method (e). However, k is not screening method-specific, i.e., k does not depend on the screening program s_j. As a result, the program that offers the lowest cost-utility ratio, i.e., the one that CUA would recommend to be implemented, is the program that offers the lowest CS-ratio. ■

Theorem 2 *Under Assumptions 1* and 2*, s is CUA superior to $\widehat{s}$ if and only if the cost per impaired individual associated to s is lower than the cost per impaired individual associated to $\widehat{s}$.*

Proof.

Let d be a particular disease and $s = s(se, sp, c)$ be a particular screening program for the early detection of d. Let $\widehat{s}$ be the screening program s but only applied to the population $\Gamma^R \subset \Gamma$ of individual with risk factors of suffering d. Let r be the probability of being in Γ^R and p_r the prevalence of the disease in Γ^R. Given a QALY index Q, the expected number of QALYs for an individual in the status quo is:

$$Q^0 = p \cdot Q_d + (1 - p) \cdot Q_{nd}.$$

The expected number of QALYs for an individual after implementing s and $\widehat{s}$ are

$$Q^s = p \cdot se \cdot Q^s_{TP} + (1-p) \cdot (1-sp) \cdot Q^s_{FP} + p \cdot (1-se) \cdot Q^s_{FN} + (1-p) \cdot sp \cdot Q^s_{TN},$$

and

$$\begin{aligned} Q^{\widehat{s}} &= r \cdot p_r \cdot se \cdot Q^s_{TP} + r \cdot (1-p_r) \cdot (1-sp) \cdot Q^s_{FP} + \\ & \quad r \cdot p_r \cdot (1-se) \cdot Q^s_{FN} + r \cdot (1-p_r) \cdot sp \cdot Q^s_{TN} + \\ & \quad (p - r \cdot p_r) \cdot Q_d + [(1-r) - (p - r \cdot p_r)] \cdot Q_{nd}, \end{aligned}$$

respectively. Thus,

$$\begin{aligned} Q^s - Q^0 &= p \cdot se \cdot Q^s_{TP} + (1-p) \cdot (1-sp) \cdot Q^s_{FP} + \\ & \quad p \cdot (1-se) \cdot Q^s_{FN} + (1-p) \cdot sp \cdot Q^s_{TN} - \\ & \quad p \cdot Q_d - (1-p) \cdot Q_{nd}, \end{aligned}$$

or equivalently,

$$Q^s - Q^0 = p \cdot (se \cdot (Q^s_{TP} - Q^s_{FN}) + (Q^s_{FN} - Q_d)) + (1-p) \cdot (sp \cdot (Q^s_{TN} - Q^s_{FP}) + (Q^s_{FP} - Q_{nd})).$$

Furthermore,

$$\begin{aligned} Q^{\widehat{s}} - Q^0 &= r \cdot p_r \cdot se \cdot Q^s_{TP} + r \cdot (1-p_r) \cdot (1-sp) \cdot Q^s_{FP} + \\ & \quad r \cdot p_r \cdot (1-se) \cdot Q^s_{FN} + r \cdot (1-p_r) \cdot sp \cdot Q^s_{TN} + \\ & \quad (p - r \cdot p_r) \cdot Q_d + [(1-r) - (p - r \cdot p_r)] \cdot Q_{nd} - \\ & \quad p \cdot Q_d - (1-p) \cdot Q_{nd}, \end{aligned}$$

or equivalently,

$$Q^{\widehat{s}} - Q^0 = r \cdot p_r \cdot (se \cdot (Q^s_{TP} - Q^s_{FN}) + (Q^s_{FN} - Q_d)) + r \cdot (1-p_r) \cdot (sp \cdot (Q^s_{TN} - Q^s_{FP}) + (Q^s_{FP} - Q_{nd})).$$

By Assumptions 1* and 2* we have

$$Q^s - Q^0 = p \cdot \gamma_Q \cdot se,$$

and

$$Q^{\widehat{s}} - Q^0 = r \cdot p_r \cdot \gamma_Q \cdot se,$$

where $\gamma_Q = Q^s_{TP} - Q^s_{FN} > 0$. Thus, the cost-utility ratios of s and $\widehat{s}$ are

$$\frac{c}{p \cdot \gamma_Q \cdot se} \text{ and } \frac{\widehat{c}}{r \cdot p_r \cdot \gamma_Q \cdot se},$$

respectively. Consequently, s is CUA superior to $\widehat{s}$ if and only if

$$\frac{c}{p \cdot \lambda_Q \cdot se} \leq \frac{\widehat{c}}{r \cdot p_r \cdot \lambda_Q \cdot se},$$

or what is equivalent, if and only if,

$$\frac{c}{p} \leq \frac{\widehat{c}}{r \cdot p_r}.$$

In other words, s is CUA superior to $\widehat{s}$ if and only if the cost per impaired individual associated to s is lower than the cost per impaired individual associated to $\widehat{s}$. ■

References

[1] Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes. Second Edition*. Oxford University Press. Oxford, U.K. 1997.

[2] Read JL, Quinn RJ, Berwick DM, Finebeg HV, Weinstein MC. *Preferences for health outcomes: comparison of assessment methods.* Med Decis Making. 1984;4:315-29.

[3] Bleichrodt H, Johannesson M. *An experimental test of the theoretical foundation for rating-scale valuations*. Med Decis Making. 1997;17:208-16

[4] Bleichrodt H, Johannesson M. *Standard Gamble, Time Trade-off and Rating Scale: Experimental Results on the Ranking Properties of QALYs*. J Health Econ. 1997;16:155-75.

[5] Bleichrodt H. *A New Explanation for the Difference Between SG and TTO Utilities*. Health Econ. 2002;11:447-56.

[6] Gold, MR, Siegel, JE, Rusell, LB, Weinstein, MC. *Cost-effectiveness in health and medicine.* Oxford University Press. New York. 1996.

[7] Bleichrodt H, Gafni A. *Time preference, the discounted utility model and health care*. J Health Econ. 1996;15:49-67.

[8] Cairns J, van der Pol M. *Negative and zero time preference for health.* Health Econ. 2000;9:171-5.

[9] National Institutes of Health. *Consensus statement. Early identification of hearing impairment in infants and young children.* NIH Consensus Statement. Mar 1-3 1993;11**:**1-24.

[10] European consensus project on neonatal hearing screening. *European consensus statement on neonatal hearing screening.* Eur. J. Pediatr. 1999;158:95-6.

[11] Joint Committee on Infant Hearing. *Year 2000 position statement: principles and guidelines for early hearing detection and intervention programs.* Am. J. Audiol. 2000;9:9-29.

[12] Keren R, Helfand M, Homer CJ, McPhillips H, Lieu TL, (2002) *Projected cost-effectiveness of statewide universal newborn hearing screening: summary of evidence.* Pediatrics; 110(5). 855-64.

[13] Sweets JA. *ROC Analysis applied to the evaluation of diagnostic techniques.* Investigative Radiology 1979;14:109-21.

[14] Sweets JA. *Measuring the accuracy of diagnostic systems.* Science 1988;240(4857):1285-93.

[15] Bess FH, Paradise JL. *Universal screening for infant hearing impairment: not simple, not risk-free, not necessarily beneficial and not presently justified*. Pediatrics. 1994;93(2): 330-4.

[16] Downs MP, Yoshinaga-Itano C. *The efficacy of early identification and intervention for children with hearing impairment.* Pediatr. Clin. North. Am. 1999;46 (1):79-87.

[17] Thompson DC, McPhillips H, Davis RL, Lieu TL, Homer CJ, Helfand M. (2001) *Universal newborn hearing screening: summary of evidence.* JAMA. Oct 24-31; 286(16). 2000-10.

[18] Kemper AR, Downs SM. *A cost-effectiveness analysis of newborn hearing screening strategies*. Arch. Pediatr. Adolesc. Med. 2000;154(5):484-8.

[19] Kezirian EJ, White KR, Yueh B, Sullivan S. *Cost and cost-effectiveness of universal screening for hearing loss in newborns*. Otolaryngol Head Neck Surg. 2001;124(4):359-67.

[20] Herrero C, Moreno-Ternero JD. *Economic evaluation of newborn hearing screening procedures*. IVIE Working Paper "A Discusión". 2002;WP-AD 2002-06.

[21] Herrero C, Moreno-Ternero JD. *Production gains associated to screening programs: an equality of opportunity approach*. **Mimeo.** 2003. Universidad de Alicante.

[22] Clemens CJ, Davis SA, Bailey AR. *The false positive in universal newborn hearing screening.* Pediatrics. 2000;106(1):E7.

[23] Stuart A, Moretz M, Yang E. *An investigation of maternal stress after neonatal hearing screening.* Am. J. Audiol. 2000;9:1-7.

[24] Weichbold V, Welzl-Mueller K, Mussbacher E. *The impact of information on maternal attitudes towards universal neonatal hearing screening.* Br. J. Audiol. 2001;35(1):59-66

[25] Weichbold V, Welzl-Mueller K. *Maternal concern about positive test results in universal newborn hearing screening.* Pediatrics 2001;108(5):1111-6.

[26] Laupacis A, Feeny D, Detsky AS, Tugwell PX *How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations.* CMAJ. 1992 Feb 15;146(4):473-81.

[27] Briggs A, Gray A. *Using cost effectiveness information.* BMJ 2000;320:246